April 2008

Further Results and Analysis of Icelandic Part of Speech Tagging

Mark Dredze
University of Pennsylvania, mdredze@cis.upenn.edu

Joel Wallenberg
University of Pennsylvania, joel.wallenberg@gmail.com

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-08-13.

Further Results and Analysis of Icelandic Part of Speech Tagging

Abstract
Data driven POS tagging has achieved good performance for English, but can still lag behind linguistic rule based taggers for morphologically complex languages, such as Icelandic. We extend a statistical tagger to handle fine grained tagsets and improve over the best Icelandic POS tagger. Additionally, we develop a case tagger for non-local case and gender decisions. An error analysis of our system suggests future directions. This paper presents further results and analysis to the original work (Dredze and Wallenberg, 2008).

Comments
University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-08-13.

Further Results and Analysis of Icelandic Part of Speech Tagging

Mark Dredze
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104
mdredze@cis.upenn.edu

Joel Wallenberg
Department of Linguistics
University of Pennsylvania
Philadelphia, PA 19104
joelcw@babel.ling.upenn.edu

Abstract

Data driven POS tagging has achieved good performance for English, but can still lag behind linguistic rule based taggers for morphologically complex languages, such as Icelandic. We extend a statistical tagger to handle fine grained tagsets and improve over the best Icelandic POS tagger. Additionally, we develop a case tagger for non-local case and gender decisions. An error analysis of our system suggests future directions. This paper presents further results and analysis to the original work (Dredze and Wallenberg, 2008).

Technical Report MS-CIS-08-13, University of Pennsylvania, Department of Computer and Information Science, 2008.

1 Introduction

While part of speech (POS) tagging for English is very accurate, languages with richer morphology demand complex tagsets that pose problems for data driven taggers. In this work we consider Icelandic, a language for which a linguistic rule-based method is the current state of the art, indicating the difficulty this language poses to learning systems. Like Arabic and Czech, other morphologically complex languages with large tagsets, Icelandic can overwhelm a statistical tagger with ambiguity and data sparsity.

Shen et al. (2007) presented a new framework for bidirectional sequence classification that achieved the best POS score for English. In this work, we evaluate their tagger on Icelandic and improve results with extensions for fine grained annotations. Additionally, we show that good performance can be achieved using a strictly data-driven learning approach without external linguistic resources (morphological analyzer, lexicons, etc.). Our system achieves the best performance to date on Icelandic, with insights that may help improve other morphologically rich languages.

After some related work, we describe Icelandic morphology followed by a review of previous approaches. We then apply a bidirectional tagger and extend it for fine grained languages. A tagger for case further improves results. We conclude with an analysis of remaining errors and challenges.

2 Related Work

Previous approaches to tagging morphologically complex languages with fine grained tagsets have considered Czech and Arabic. Khoja (2001) first introduced a tagger for Arabic, which has 131 tags, but subsequent work has collapsed the tagset to simplify tagging (Diab et al., 2004). Like previous Icelandic work (Loftsson, 2006), morphological analyzers disambiguate words before statistical tagging in Arabic (Habash and Rambow, 2005) and Czech (Hajič and Hladká, 1998). This general approach has led to the serial combination of rule based and statistical taggers for efficiency and accuracy (Hajič et al., 2001). While our tagger could be combined with these linguistic resources as well, as in Loftsson (2007b), we show state of the art performance without these resources. Another fine-grained tagging approach captures grammatical structures with tree-based tags, such as "supertags" in the tree-adjoining grammar of Bangalore and Joshi (1999).

3 Icelandic Morphology

Icelandic is a North Germanic language spoken by approximately 313,000 people, and is notable for its morphological richness as well as a number of interesting syntactic properties (see Þráinsson (1994) for an overview). Verbs can show different morphology depending on Tense, Mood (subjunctive or indicative), Voice (Middle or non-Middle), Person, and Number; these properties potentially yield 54 different combinations. A highly productive class of verbs also show stem vowel alternations reminiscent Semitic verb morphology (Arabic). Noun morphology exhibits a robust case system; nouns may appear in as many as 16 different forms, depending on Number, Definiteness (presence or absence of a definite article suffix), and Case (nominative, accusative, dative, or genitive).

The four-case system of Icelandic is similar to that of the Slavic languages (Czech), with case morphology also appearing on elements which agree in case with nouns, i.e. determiners, adjectives, past participles. However, unlike Czech, case frequently does not convey distinct meaning in Icelandic as it is often determined by elements such as the governing verb in a clause (non-local information). Therefore, while Icelandic case looks formally like Slavic and presents similar challenges for POS tagging, it also may be syntactically-determined, as in Standard Arabic.

Word order in Icelandic is quite constrained compared to that of Czech and other case-rich languages, though Icelandic syntax is still somewhat freer than that of a language like English. It is a verb-second (V2) language, and so although the most common order of elements in a clause is the English-like order of Subject-Verb-Object, OVS and XVSO orders are also possible. Icelandic also allows a very limited form of scrambling, "object shift", which permutes the order of objects and some sentential modifiers, but does not produce anything like the variety of word-order permutations allowed in the Slavic languages. This combination of morphological complexity and syntactic constraint makes Icelandic a good case study for statistical POS tagging techniques.

The result of this complexity is the very large extended tagset developed for the Icelandic Frequency Dictionary (Íslensk orðtíðnibók/IFD), a corpus of roughly 590,000 tokens (Pind et al., 1991). We use the 10 IFD data splits produced by Helgadóttir (2004), where the first nine splits are used for evaluation and the tenth for model development. Tags are comprised of up to six elements, such as word class, gender, number, and case, yielding a total of 639 tags, not all of which occur in the training data. This is far more than the 131 tags in Arabic (Khoja, 2001) but not as many as the 3127 in Czech ((Hajič and Hladká, 1998)). A prediction is correct when an entire tag matches; partial matches are counted as incorrect.

4 Previous Approaches

Helgadóttir (2004) evaluated several data-driven models for Icelandic, including MXPost, a maximum entropy tagger, and TnT, a trigram HMM, as well as several transformational based and memory-based learning algorithms. Both MXPost and TnT did considerably worse than on English. Icelandic poses significant challenges: data sparseness, non-local tag dependencies, and 136,264 observed trigram sequences make discriminative sequence models, such as CRFs, prohibitively expensive.

Given these challenges, the most successful tagger is IceTagger (Loftsson, 2007b), a linguistic rule based system with several linguistic resources: a morphological analyzer, a series of local rules and heuristics for handling PPs, verbs, and forcing agreement. IceTagger processes a sentence in several passes. First, it introduces ambiguity by creating a list of potential tags for each word based on an analysis of the training data and a morphological analyzer. Next, a disambiguation pass eliminates many of the potential tags using 175 local rules and a number of global heuristics to enforce agreement. If multiple tags remain the most commonly used tag is selected. The final system achieves an 11.5% reduction in error over TnT. For a detailed explanation of IceTagger see Loftsson (2007a)[Chapter 5].

Loftsson also improves the best statistical tagger (TnT) by integrating a morphological analyzer (TnT*). The analyzer is used to reduce the number of tags considered by TnT, which in turn leads to an .74% increase in performance. Helgadóttir (2004) considers other ways of combining statistical sys-

það	er	auðveldara	en	margt	annað
fphen	sfg3en	lhenvm	c	lhensf	fohen
it	is	simpler	than	many	another

Figure 1: “It is simpler than many others.”

tems together, such as classifier combination, which yields the best combined system score for Icelandic.

Despite the challenges of a purely statistical system, data driven taggers have several advantages. Learning systems can be easily applied to new corpora, tagsets, or languages and can accommodate integration of other systems (including rule based) or new linguistic resources, such as those used by Loftsson. Therefore, we seek a learning system that can handle these challenges.

5 Bidirectional Sequence Classification

Bidirectional POS tagging (Shen et al., 2007), the current state of the art for English, has some properties that make it appropriate for Icelandic. For example, it can be trained quickly with online learning and does not use tag trigrams, which reduces data sparsity and the cost of learning. It can also allow long range dependencies, which we consider below.

Bidirectional classification uses a perceptron style classifier to assign potential POS tags (hypotheses) to each word using standard POS features and some additional local context features. On each round, the algorithm selects the highest scoring hypothesis and assigns the guessed tag. Unassigned words in the context are reevaluated with this new information. If an incorrect hypothesis is selected during training, the algorithm promotes the score of the correct hypothesis and demotes the selected one. See Shen *et al.* for a detailed explanation.

Consider an example from the IFD corpus, shown with translation and tags in figure 1. A guide to some of the tags is in figure 2. The tagger begins by selecting the most likely tag for each word, and the applies the single highest scoring tag across all words. The tagger first selects the verb “er”, a common verb in the corpus.

það	er	auðveldara	en	margt	annað
	sfg3en				

Next, the tagger uses the verb tag as a context feature to identify “það” as a nominative pronoun and then “auðveldara” as a nominative adjective.

það	er	auðveldara	en	margt	annað
fphen	sfg3en	lhenvm			

The tagger continues in this fashion until the entire sentence has been tagged.

We begin with a direct application of the bidirectional tagger to Icelandic using a beam of one and the same parameters and features as Shen *et al.* [1] On the development split the tagger achieved an accuracy of 91.61%, which is competitive with the best Icelandic systems. However, test evaluation is not possible due to the prohibitive cost of training the tagger on nine splits; training took almost 4 days on an AMD Opteron 2.8 GHz machine.

Tagset size poses a problem since the tagger must evaluate over 600 options to select the top tag for a word. The tagger rescores the local context after a tag is committed or all untagged words if the classifier is updated. This also highlights a problem with the learning model itself. The tagger uses a one vs. all multi-class strategy, requiring a correct tag to have higher score than every other tag to be selected. While this is plausible for a small number of labels, it overly constrains an Icelandic tagger.

As with most languages, it is relatively simple to assign word class (noun, verb, etc.) and we use this property to divide the tagset into separate learning problems. First, the tagger classifies a word according to one of the eleven word classes. Next, it selects and evaluates all tags consistent with that class. When an incorrect selection is updated, the word class classifier is updated only if it was mistaken as well. The result is a dramatic reduction in the number of tags considered at each step. For some languages, it may make sense to consider further reductions, but not for Icelandic since case, gender, and number decisions are interdependent. Additionally, by learning word class and tag separately, a correct tag need only score higher than other tags of the same word class, not all 639. Furthermore, collapsing tags into word class groups increases train-

[1] While we tried no parameter optimization, tests with a beam size of 3 yielded no improvements.

Char #	*Category/Feature*	*Symbol – semantics*
1	Word Class	**f**-pronoun
2	Subcategory	**p**-personal, **o**-indefinite
3	Gender/Person	**k**-masculine, **v**-feminine, **h**-neuter
4	Number	**e**-singular, **f**-plural
5	Case	**n**-nominative, **o**-accusative, **þ**-dative, **e**-genitive
1	Word class	**s**-verb
2	Mood	**n**-infinitive, **f**-indicative, **v**-subjunctive
3	Voice	**g**-active, **m**-middle
4	Person	**1**-1st person, **2**-2nd person, **3**-3rd person
5	Number	**e**-singular, **f**-plural
6	Tense	**n**-present, **þ**-past
1	Word class	**l**-adjective
2	Gender	**k**-masculine, **v**-feminine, **h**-neuter
3	Number	**e**-singular, **f**-plural
4	Case	**n**-nominative, **o**-accusative, **þ**-dative, **e**-genitive
5	Declension	**s**-strong declension, **v**-weak declension, **o**-indeclineable
6	Degree	**f**-positive, **m**-comparative, **e**-superlative

Figure 2: A partial list of tags from the Icelandic tagset for figure 1.

ing data, allowing the model to generalize features over all tags in a class instead of learning each tag separately (a form of parameter tying).

Training time dropped to 12 hours with the bidirectional word class (WC) tagger and learning performance increased to 91.98% (table 1). [2] The tagger learned an average of 5,841,526 features per training split. Word class accuracy, already quite high at 97.98%, increased to 98.34%, indicating that the tagger can quickly filter out most inappropriate tags. The reduced training cost allowed for test data evaluation, yielding 91.68%, which is a 12.97% relative reduction in error over the best pure data driven model (TnT) and a 1.65% reduction over the best model (IceTagger).

	Accuracy			*Train*
Tagger	*All*	*Known*	*Unkn.*	*Time*
Bidir	91.61	93.21	69.76	90:27
Bidir+WC	91.98	93.58	70.10	12:20
Bidir+WC+CT	92.36	93.93	70.95	14:02

Table 1: Results on development data. Accuracy is measured by exact match with the gold tag. About 7% of tokens are unknown at test time.

Tagger	*Unknown*	*Known*	*All*
MXPost	62.50	91.04	89.08
TnT	71.68	91.82	90.44
TnT*	72.75	92.53	91.18
IceTagger	75.09	92.74	91.54
Bidir+WC	69.25	93.32	91.68
Bidir+WC+CT	69.74	93.70	92.06

Table 2: Results on test data.

6 Case Tagger

Examining tagger error reveals that most mistakes are caused by case confusion on nouns (84.61% accuracy), adjectives (76.03%), and pronouns (90.67%); these account for 40% of the corpus. Table 3 shows a breakdown of accuracy by case and gender for nouns, adjectives and pronouns. As we noted in section 3, there are 16 case-number-definiteness combinations in Icelandic noun morphology, and so there are potentially 16 different forms for each noun. However, depending on the gender and noun class, a number of these combinations may actually be realized by a single phonological/orthographic form (case-syncretism). When the

[2] We note that prediction time is still slower than other systems. Loftsson (2007b) reports a speed of 2,700 tokens/sec for IceTagger. Our tagger processed 179 tokens/sec.

Word Class	*Feature*	*Accuracy*	*CT Accuracy*
Adjective	Case	90.95	93.09
	Gender	92.10	92.10
Noun	Case	90.37	92.11
	Gender	95.87	95.31
Pronoun	Case	96.75	97.15
	Gender	95.79	95.79

Table 3: Accuracy on case and gender for nouns, adjectives and pronouns on the test data.

different case forms are distinct, the best clue to the case of a noun is the form of the noun. However, for some nouns, 2–4 different case possibilities may all be plausible by looking at the word itself. Mistakes in noun case lead to further mistakes for categories which agree with nouns, e.g. adjectives. Assigning appropriate case for nouns is an important step in making a number of other tagging decisions, such as identifying the finite verb of a clause and identifying adjectives. However, the case on a noun often provides little or no information about the identity of other tags. It is in this situation that the tagger makes most case-assignment errors. Therefore, while accuracy depends on correct case assignment for these nouns, other tags are mostly unaffected.

One approach to correcting these errors is to introduce long range dependencies, such as those used by IceTagger. While normally hard to add to a learning system, bidirectional learning provides a natural framework since non-local features can be added once a tag has been committed. To allow dependencies on all other tag assignments, and because correcting the remaining case assignments is unlikely to improve other tags, we constructed a separate bidirectional case tagger (CT) that retags case on nouns, adjectives and pronouns. [3] This approach is motivationally similar to transformation-based learning (Brill, 1995). Since gender is important as it relates to case, it is retagged as well. The case tagger takes a fully tagged sentence from the POS tagger and retags case and gender to nouns, adjectives and pronouns.

[3]We considered adding case tagging features to and removing case decisions from the tagger; both hurt performance.

6.1 Case Tagging Features

The case tagger uses the same features as the POS tagger, but it now has access to all predicted tags. Additionally, we develop several non-local features. We describe each of these new feature types and evaluate their performance for case tagging on development data. The best features are selected for test performance.

Many case decisions are entirely idiosyncratic, even from the point of view of human language-learners. Some simple transitive verbs in Icelandic arbitrarily require their objects to appear in dative or genitive case, rather than the usual accusative. This arbitrary case-assignment adds no additional meaning, and this set of idiosyncratic verbs is memorized by speakers. A statistical tagger likewise must memorize these verbs based on examples in the training data. A similar problem is presented by a loosely semantically-related class of verbs (experiencer verbs) which idiosyncratically take subjects that are dative or accusative, rather than default case for subjects, nominative (the quirky case phenomenon). While there is some evidence that speakers can learn this group of verbs as a class based on their meanings, which are related in some ways, it is unlikely that a statistical tagger could do so with a large amount of training data.

To aid generalization, verb-forms were augmented by verb-stem-features. A simple regular expression generated stem candidates: the first sequence of `j` plus one or two vowels was replaced with a wildcard (to generalize over the vowel changes that some verb roots undergo in different tenses) and the 2-3 final characters were removed for past tense or past participles (as long as at least 2 characters remained) or 1-2 final characters for present tense verbs to abstract away from person-number-tense suffixes. While crude, it frequently yielded accurate stems: e.g., the verb forms `dveldi`, `dvaldi`, `dvelst`, `dvelur` all mapped to the stem `dv*l` (*dvelja* "dwell"). Stems that occurred less than three times in the training data were ignored. Additional discovered stems are shown in table 4.

Using these stems we created several groups of features.

- **Feature Group 1** - For each word, features

Stem	*Verbs*
dr*kk	drekkur, drekkum, drukkið, drukkna drukkum, drekktu, drykkju
læs	læsir, læsti, læstum
r*nd	reyndust, reyndist, reyndu, reyndir reyndum
h*gs	hugsar, hugsuðu, hugsi, hugsaði, hugsum hugsa, hugsað, hugsið
*rð	orðnar, orðað, orða, orðnir, urðum, orðar orðin, orðuðu, yrðir, orðaði, orðið, orðinn yrðum, orðuð

Table 4: Stems and their verbs discovered by the stemming rule.

Features	*Accuracy*
POS Tagger feature only	92.25
Feature Group 1	**92.38**
Feature Group 2	**92.28**
Feature Group 3	92.24
Feature Group 4	**92.26**
Feature Group 5	92.21
Feature Group 6	92.17
Feature Group 7	92.22
Stem features (1+2+4)	92.36

Table 5: Accuracy on development data using different features for the case tagger. Runs that improved over the baseline are in bold.

were added indicating the preceding verb's (predicted) tag, gender, case, stem, verb, and the verb with the last 2 and 3 characters removed.

- **Feature Group 2** - If the word was closer to the verb than any noun (in either direction), this was conjoined with the verb's tag, case, and stem. Other features that used noun position: the number of preceding nouns and the number of nouns since the last clause boundary, indicated by the tags `cn` (complementizer) or `ct` (relativizer).

- **Feature Group 3** - The stem of any verbs that occurred to the immediate left and right of this word.

- **Feature Group 4** - The stem, direction and distance of verbs were conjoined if they occurred within 2 words.

- **Feature Group 5** - Find the previous preposition, numeral and article that occurred within 5 words of the current word. Add their case, tag, and gender.

- **Feature Group 6** - The number of nouns that occur before this word (0, 1, or more). If one noun occurs before, then its case, tag and gender. If more than one noun occurs before this word, both of their cases, tags, and genders. A feature to indicate if this word is before or after the first verb in the sentence and a conjunctions with its tag, case and gender.

- **Feature Group 7** - Select the closest nouns and pronouns occurring to the left and right of the current word within 5 words. Create features indicating the direction and current predicted tag for the noun/pronoun. Additional features for the direction and word class and, if assigned, case.

Each feature was added to the baseline feature set of the POS tagger and evaluated on the development data. Results are shown in table 5. The feature groups that contained stem information improved the case tagger's performance and we selected these as the features for the case tagger.

6.2 Test Results

The case tagger was used to correct the output of the tagger after training on the corresponding train split. Case tagging gold input (all other decisions correct) yielded a total tag accuracy of 97.48%, indicating the case tagger can effectively make difficult case decisions. The case tagger improved results yielding a new best accuracy of 92.06%, a 16.95% and 12.53% reduction over the best data driven and rule systems. Results showing improvements in case and gender are shown in table 3.

To illustrate the effect of case tagging, we return to our previous example sentence (figure 1). The

Word Class	*Accuracy*	*Total Tokens*
Adjective	77.57	35,669
Adverb/Preposition	94.97	116,112
Article	81.33	632
Conjunction	98.29	60,256
Foreign Word	32.36	411
Noun	85.77	122,621
Numeral	81.29	5901
Pronoun	91.04	74,315
Unanalyzed	37.80	127
Verb	94.04	103,136

Table 6: Breakdown of final accuracy and total tokens by word class on test data.

POS tagger correctly tagged this sentence except for the case of the final adjective and pronoun (margt annað), which were incorrectly assigned accusative and not dative case. The case tagger was able to learn using non-local features that these forms were actually dative case, and the tags were corrected.

7 Remaining Challenges

We have shown that a data driven approach can achieve state of the art performance on highly inflected languages by extending bidirectional learning to fine grained tagsets and designing a bidirectional non-local case tagger. We conclude with an error analysis to provide future direction.

The tagger is particularly weak on unknown words, a problem caused by case-syncretism and idiosyncratic case-assignment. Data driven taggers can only learn which verbs assign special object cases by observation in the training data. Some verbs and prepositions also assign case based on the meaning of the whole phrase. These are both serious challenges for data-driven methods and could be addressed with the integration of linguistic resources.

However, there is more work to be done on data driven methods. Mistakes in case-assignment due to case syncretism, especially in conjunction with idiosyncratic-case-assigning verbs, account for a large proportion of remaining errors. Verbs that take dative rather than accusative objects are a particular problem, such as mistaking accusative for dative feminine objects (10.6% of occurrences) or dative for accusative feminine objects (11.9%). Table 6 shows accuracy by word class and the total tokens of that class in the corpus. Table 7 lists the most common mistaggings. A possible learning solution lies in combining POS tagging with syntactic parsing, allowing for the identification of clause boundaries, which may help disambiguate noun cases by deducing their grammatical function from that of other clausal constituents.

Additionally, idiosyncratic case-assignment could be learned from *unlabeled* data by finding unambiguous dative objects to identify idiosyncratic verbs. Furthermore, our tagger learns which prepositions idiosyncratically assign a single odd case (e.g. genitive) since prepositions are a smaller class and appear frequently in the corpus. This indicates that further work on data driven methods may still improve the state of the art.

8 Acknowledgments

We thank Hrafn Loftsson for sharing IceTagger and the datasplits, Libin Shen for his tagger, and the Árni Magnússon Institute for Icelandic Studies for access to the corpus.

References

Srinivas Bangalore and Arivand K. Joshi. 1999. Supertagging: An approach to almost parsing. *Computational Linguistics*, 25(2).

Eric Brill. 1995. Transformation-based error-driven learning and natural language processing: A case study in part of speech tagging. *Computational Linguistics*.

Mona Diab, Kadri Hacioglu, and Daniel Jurafsky. 2004. Automatic tagging of arabic text: From raw text to base phrase chunks. In *NAACL/HLT*.

Mark Dredze and Joel Wallenberg. 2008. Icelandic data driven part of speech tagging. In *ACL*.

Nizar Habash and Owen Rambow. 2005. Arabic tokenization, part-of-speech tagging and morphological disambiguation in one fell swoop. In *ACL*.

Jan Hajič and Barbora Hladká. 1998. Tagging inflective languages: prediction of morphological categories for a rich, structured tagset. In *COLING*.

Jan Hajič, Pavel Krbec, Pavel Květoň, Karel Oliva, and Vladimír Petkevič. 2001. Serial combination of rules and statistics: a case study in czech tagging. In *ACL*.

Sigrun Helgadóttir. 2004. Testing data-driven learning algorithms for pos tagging of icelandic.

Shereen Khoja. 2001. Apt: Arabic part-of-speech tagger. In *NAACL Student Workshop*.

Gold Tag	*Predicted Tag*	*Mistake Rate*	*Tag Count*	*Mistake*
nveo	nveþ	10.59%	5478	Dative for accusative
nkeo	nkeþ	5.41%	5193	Dative for accusative
nven	nveo	3.15%	5081	Accusative for nominative
nveþ	nveo	11.96%	5026	Accusative for dative
nkeþ	nkeo	7.48%	5012	Accusative for dative
nheo	nhen	5.97%	4877	Nominative for accusative
lhensf	aa	5.03%	4119	Preposition for adjective
sfg3fn	sng	13.77%	3683	Infinitive for indicative
svg3eþ	sfg3eþ	7.84%	3609	Indicative for subjunctive
nhen	nheo	8.99%	3350	Accusative for nominative
sfg1eþ	sfg3eþ	16.73%	2767	Third for first person
sþghen	ssg	10.58%	2713	Supine for past participle
nvfo	nvfn	8.36%	2477	Nominative for accusative
nhfo	nheo	12.44%	2251	Singular for plural
nvfn	nvfo	16.30%	1822	Accusative for nominative
nhfn	nhfo	12.73%	1697	Accusative for nominative
fpkeo	fpken	8.61%	1661	Nominative for accusative
nheng	nheog	8.48%	1616	Accusative for nominative
fpheo	fphen	9.89%	1567	Nominative for accusative
fpveþ	fpkeþ	19.93%	1525	Masculine for feminine

Table 7: Common tagging mistakes.

Hrafn Loftsson. 2006. Tagging icelandic text: A linguistic rule-based approach. Technical Report CS-06-04, University of Sheffield.

Hrafn Loftsson. 2007a. *Tagging and Parsing Icelandic Text*. Ph.D. thesis, University of Sheffield, UK, May.

Hrafn Loftsson. 2007b. Tagging icelandic text using a linguistic and a statistical tagger. In *NAACL/HLT*.

J Pind, F Magnússon, and S Briem. 1991. The icelandic frequency dictionary. Technical report, The Institute of Lexicography, University of Iceland.

Libin Shen, Giorgio Satta, and Aravind K. Joshi. 2007. Guided learning for bidirectional sequence classification. In *ACL*.

Höskuldur Þráinsson. 1994. Icelandic. In Ekkehard König and Johan van der Auwera, editors, *The Germanic Languages*. Routledge, London.